



Multivoxel Patterns Reveal Functionally Differentiated Networks Underlying Auditory Feedback Processing of Speech

Zheng, Zane Z.; Vicente-Grabovetsky, Alejandro; MacDonald, Ewen N.; Munhall, Kevin G.; Cusack, Rhodri; Johnsrude, Ingrid S.

Published in:
Journal of Neuroscience

Link to article, DOI:
[10.1523/JNEUROSCI.6319-11.2013](https://doi.org/10.1523/JNEUROSCI.6319-11.2013)

Publication date:
2013

Document Version
Publisher's PDF, also known as Version of record

[Link back to DTU Orbit](#)

Citation (APA):
Zheng, Z. Z., Vicente-Grabovetsky, A., MacDonald, E. N., Munhall, K. G., Cusack, R., & Johnsrude, I. S. (2013). Multivoxel Patterns Reveal Functionally Differentiated Networks Underlying Auditory Feedback Processing of Speech. *Journal of Neuroscience*, 33(10), 4339-4348. <https://doi.org/10.1523/JNEUROSCI.6319-11.2013>

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Multivoxel Patterns Reveal Functionally Differentiated Networks Underlying Auditory Feedback Processing of Speech

Zane Z. Zheng,¹ Alejandro Vicente-Grabovetsky,⁴ Ewen N. MacDonald,^{2,5} Kevin G. Munhall,^{1,2,3} Rhodri Cusack,^{4,6,7,8} and Ingrid S. Johnsrude^{1,2}

¹Centre for Neuroscience Studies, ²Department of Psychology, and ³Department of Otolaryngology, Queen's University, Kingston, K7L 3N6 Ontario, Canada, ⁴MRC Cognition and Brain Sciences Unit, Cambridge, CB2 7EF, United Kingdom, ⁵Centre for Applied Hearing Research, Technical University of Denmark, 2800 Kgs. Lyngby, Denmark, and ⁶Centre for Brain and Mind, ⁷Department of Psychology, and ⁸Department of Medical Biophysics, University of Western Ontario, N6A 5B7 Ontario, Canada

The everyday act of speaking involves the complex processes of speech motor control. An important component of control is monitoring, detection, and processing of errors when auditory feedback does not correspond to the intended motor gesture. Here we show, using fMRI and converging operations within a multivoxel pattern analysis framework, that this sensorimotor process is supported by functionally differentiated brain networks. During scanning, a real-time speech-tracking system was used to deliver two acoustically different types of distorted auditory feedback or unaltered feedback while human participants were vocalizing monosyllabic words, and to present the same auditory stimuli while participants were passively listening. Whole-brain analysis of neural-pattern similarity revealed three functional networks that were differentially sensitive to distorted auditory feedback during vocalization, compared with during passive listening. One network of regions appears to encode an “error signal” regardless of acoustic features of the error: this network, including right angular gyrus, right supplementary motor area, and bilateral cerebellum, yielded consistent neural patterns across acoustically different, distorted feedback types, only during articulation (not during passive listening). In contrast, a frontotemporal network appears sensitive to the speech features of auditory stimuli during passive listening; this preference for speech features was diminished when the same stimuli were presented as auditory concomitants of vocalization. A third network, showing a distinct functional pattern from the other two, appears to capture aspects of both neural response profiles. Together, our findings suggest that auditory feedback processing during speech motor control may rely on multiple, interactive, functionally differentiated neural systems.

Introduction

The articulatory movements of speech must be produced very quickly and must be precise in their execution and timing. Auditory feedback is essential for accurate speech production (e.g., Guenther et al., 2006; Hickok et al., 2011; Houde and Nagarajan, 2011), implying a complex control system, in which speech errors are detected and processed, ultimately resulting in altered articulation. Here, with the aid of real-time acoustic feedback perturbation of produced speech and novel representational similarity

analyses of fMRI signals, we demonstrate functionally differentiated networks underlying auditory feedback processing for speech motor control.

Theoretical accounts of speech monitoring posit multiple functional components required for detection of errors in speech planning (Levelt, 1983; Levelt et al., 1999; Postma, 2000). However, neuroimaging studies generally indicate either single brain regions sensitive to speech production errors, or small, discrete networks (McGuire et al., 1996; Paus et al., 1996; Hashimoto and Sakai, 2003; Fu et al., 2006; Christoffels et al., 2007; Tourville et al., 2008; Zheng et al., 2010; Christoffels et al., 2011). The discrepancy between the complexity of theoretical accounts and neuroimaging data may be attributable to the univariate analyses that are typically conducted (but see Tourville et al., 2008 for an ROI-based analysis): these analyses are not well suited to the characterization of distributed brain networks.

Here, we use pattern-information analysis of fMRI data (Haxby et al., 2001; Haynes and Rees, 2006; Norman et al., 2006; Kriegeskorte et al., 2008; Kriegeskorte, 2011) to explore auditory feedback processing networks. Within a general linear model (GLM), we examine multivoxel neural patterns during speech production and listening with different types of auditory input,

Received Dec. 19, 2011; revised Dec. 28, 2012; accepted Jan. 25, 2013.

Author contributions: Z.Z.Z., E.N.M., K.G.M., and I.S.J. designed research; Z.Z.Z. performed research; Z.Z.Z., A.V.-G., and R.C. contributed unpublished reagents/analytic tools; Z.Z.Z., A.V.-G., E.N.M., K.G.M., and I.S.J. analyzed data; Z.Z.Z., A.V.-G., E.N.M., K.G.M., R.C., and I.S.J. wrote the paper.

This work was supported by the Canadian Institutes of Health Research (CIHR 69046) to I.S.J., the Natural Sciences and Engineering Research Council of Canada (R6PIN/327429) to I.S.J., the National Institutes of Health (R01 DC08092) to K.G.M. and I.S.J., and a Wellcome Trust Award (WT091540MA) to R.C. and A.V.-G. Z.Z.Z. was supported through a training award from the Ontario Ministry of Research and Innovation to I.S.J. I.S.J. was supported by the Canada Research Chairs program.

The authors declare no competing financial interests.

Correspondence should be addressed to either Dr. Zane Z. Zheng or Dr. Ingrid S. Johnsrude, Centre for Neuroscience Studies, Queen's University, Kingston, K7L 3N6 Ontario, Canada. E-mail: zane.z.zheng@gmail.com or ingrid.johnsrude@queensu.ca.

DOI:10.1523/JNEUROSCI.6319-11.2013

Copyright © 2013 the authors 0270-6474/13/334339-10\$15.00/0

and probe for commonalities/differences in neural response profiles across different conditions. The GLM framework provides a conceptually and computationally straightforward way to test hypotheses based on brain pattern similarity, rendering our method a simple and flexible variant within the pattern-information analysis family.

We searched the whole brain (Searchlight method, Kriegeskorte et al., 2006) for patterns of neural activity that are consistent when processing load is specifically placed on putative systems subserving the perception and processing of errors during speech production, which is essential for ongoing articulatory control. We use a real-time speech-tracking system to deliver normal feedback and two different types of distorted auditory feedback (formant-shifted speech, Houde and Jordan, 1998; and signal-correlated noise, Schroeder, 1968) in response to spoken words. These same auditory stimuli are also presented in passive-listening conditions. We look for brain regions exhibiting a distinctive neural response profile: specifically, that the two acoustically different, distorted-feedback conditions elicit similar patterns of activity, but that the patterns elicited by either of these conditions and normal feedback differ. Any such profile, present during speech production but not passive listening, can be assumed to reflect processes engaged by auditory feedback that does not match the intended motor gesture, regardless of the nature of the mismatch. Based on the neural pattern signatures, we characterize three functional networks that appear sensitive to distinct aspects of auditory feedback processing during speech motor control.

Materials and Methods

Participants. Written informed consent was obtained from 20 participants (mean age, 21 years; range, 19–27 years; 8 females). All were right handed, without any reported history of neurological or hearing disorder, and spoke English as their first language. Each participant received \$15 to compensate them for their time. Procedures were cleared by the Queen's Health Sciences Research Ethics Board.

Experimental design and stimuli. Standard behavioral paradigms to investigate auditory feedback control, from our laboratory and others (Houde and Jordan, 1998; e.g., Burnett et al., 1998; Donath et al., 2002; Purcell and Munhall, 2006a, 2006b; Villacorta et al., 2007; Munhall et al., 2009; MacDonald et al., 2010, 2011), typically involve the introduction of a feedback perturbation that changes the acoustics that participants hear and then maintain these conditions until the talker's vocalizations have measurably altered. Such designs are inappropriate for functional MRI because imposing a speech perturbation consistently over many trials (i.e., a low-frequency effect) confounds this manipulation with the slow signal fluctuations (noise) characteristic of fMRI. Furthermore, we are not specifically interested in the process of behavioral compensation (for evidence of within-utterance formant compensation using long-duration trials, see, e.g., Tourville et al. (2008)) but rather are focusing on the neural correlates of trial-specific responses to perceived error.

We used a 2×3 within-subject factorial design with 6 experimental conditions, plus a low-level silence/rest control condition. Two tasks (speech production and passive listening) were crossed with three types of auditory signal (formant-frequency-shifted speech, signal-correlated noise, and normal speech), which served as auditory stimuli during listening conditions and as temporally gated auditory feedback during production conditions. To avoid confounding the effects of our manipulation on fMRI signal with the low-frequency noise characteristic of fMRI, we adopted a paradigm in which condition types changed from trial to trial. Although we cannot measure behavioral compensation with this design, single-cell recording studies in marmosets (e.g., Eliades and Wang, 2008) and electrophysiological studies in humans (Houde et al., 2002; Heinks-Maldonado et al., 2005; Behroozmand and Larson, 2011; Behroozmand et al., 2012) show that neural responses to auditory feed-

back perturbations occur rapidly (with a latency of ~50 ms), and this would be reflected in the BOLD signal.

In the condition involving speech production with normal feedback (production-clear), participants vocalized "had" and heard unaltered auditory feedback. In the condition involving production with formant-shifted feedback (production-shift), they vocalized "head" but heard processed feedback, such that the first (F1) and second (F2) formants of the vowel were shifted by -200 Hz and 250 Hz, respectively (i.e., "head" shifted toward "hid"). The direction and magnitude of the shifts were chosen based on empirical studies of formant perturbations in our laboratory (MacDonald et al., 2010, 2011). In the condition involving production with signal-correlated noise feedback (production-noise), participants produced either "had" or "head" (pseudo-randomly cued, such that half the trials were "head" and half were "had"), and heard masking noise temporally gated with the onset and offset of their vocalizations. Having participants produce "had" half the time and "head" half the time in this condition ensured that any difference in response patterns between distorted- and normal-feedback production conditions could not simply be the result of a difference in what participants were asked to say. The three listening conditions (listen-clear, listen-shift, and listen-noise, respectively) were yoked to the production conditions, such that, on listening trials, participants heard as stimuli what they had heard as auditory feedback on earlier production trials.

Trials were presented in the 1600 ms silent period between successive 1600 ms scans. Participants were instructed to pay attention to a rectangular, gray prompt in the middle of a computer screen that appeared 100 ms before the offset of each scan, which signaled the onset of the next silent period/trial (Fig. 1). Depending upon the condition type, the prompt either contained a word or was blank. Participants were asked to produce the word when there was one, and remain silent when there was not one presented. Hence, the words "had" and "head" were shown on production-clear and production-shift trials, respectively. The words "had" or "head," presented with equal probability, were shown in the production-noise trials. The prompt was blank during listen-clear, listen-shift, listen-noise, and rest trials. Seventy-two trials of each of the seven conditions were collected during the scanning session.

Real-time speech processing. In trials of the three production conditions, participants spoke into an optical microphone (Phone-Or) and their utterances were digitized at 10 kHz with 16-bit precision using a National Instruments PXI-6052E data acquisition board (National Instruments). Real-time analysis was performed using a National Instruments PXI-8176 embedded controller. Processed acoustic signals were converted back to analog by the data acquisition board and played over high-fidelity magnet-compatible headphones (Nordic, NeuroLab) in real time. The processing delays were too short (iteration delay < 10 ms) to be noticeable to the participants. The processed signals were also recorded and stored on an IBM ThinkPad X32 laptop (IBM) to be used as yoked stimuli during listening trials.

On production-shift trials, vowel formants were estimated with an iterative Burg algorithm (Orfanidis, 1988) and formant shifting was implemented using an infinite impulse response filter in real-time (Purcell and Munhall, 2006a, 2006b). On production-noise trials, masking noise was generated by applying the amplitude envelope of the utterance to uniform Gaussian white noise. This ensures that the noise level was utterance-specific and exactly intense enough at every moment to mask the energy of the spoken words. The resulting "signal-correlated noise" (Schroeder, 1968) had the same spectral profile and amplitude envelope as the original speech signal but was completely unintelligible.

Image acquisition. fMRI data were collected on a 3-T Siemens Trio MRI system, using a rapid sparse-imaging paradigm (Orfanidou et al., 2006). This paradigm allowed us to present our auditory stimuli and record responses during a 1600 ms silent period between successive 1600 ms acquisitions to minimize acoustic interference (GE-echo planar imaging sequence; TR 3200 ms; TA 1600 ms; 211×211 matrix; in-plane resolution 3.3×3.3 mm²; 26 transverse slices with interleaved acquisition). A high-resolution T1-weighted magnetization-prepared rapid gradient echo structural scan was also acquired in each participant (TR 1760 ms, TE 2.6 ms, voxel resolution $1.0 \times 1.0 \times 1.0$ mm³, flip angle 9°).

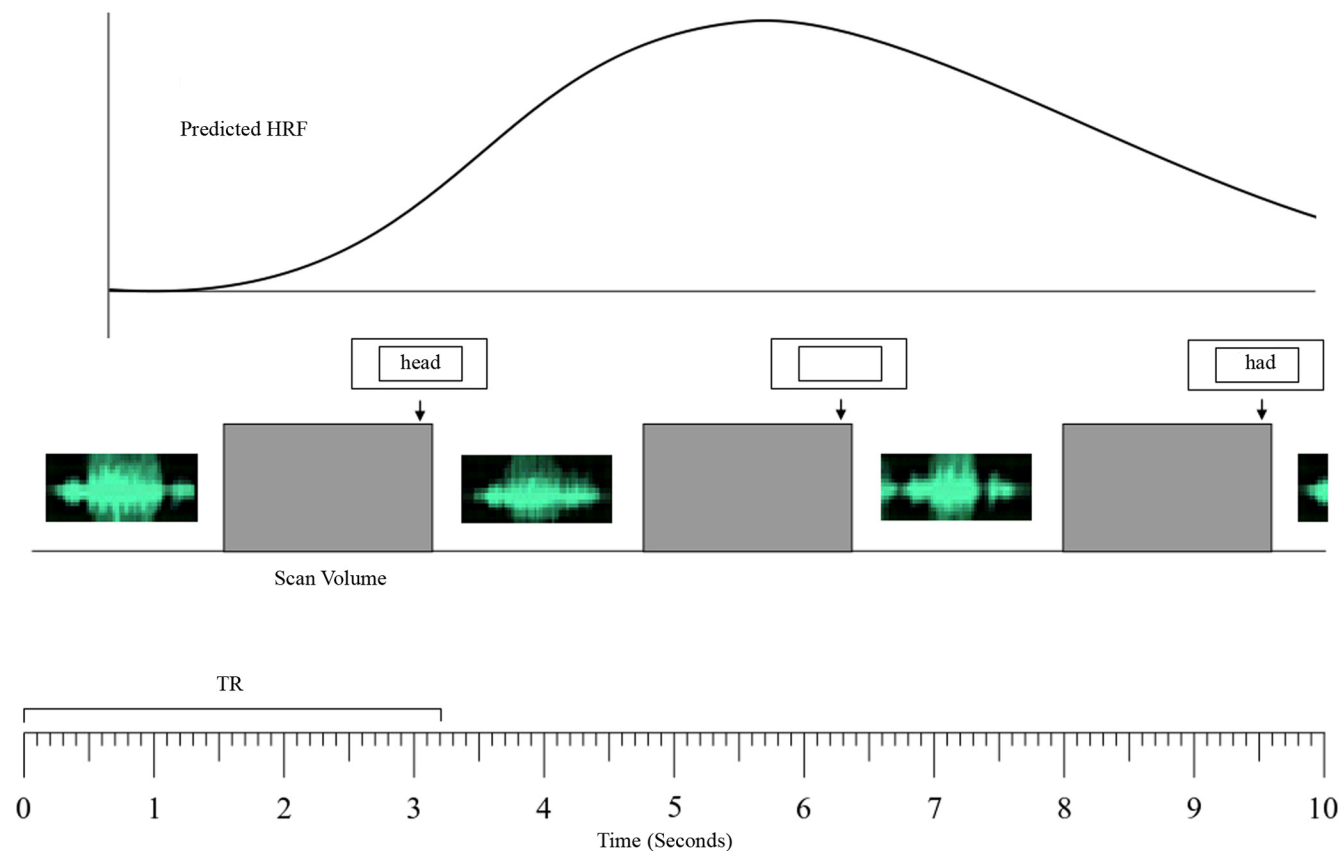


Figure 1. Schematic diagram of the first 10 s from a functional run with predicted hemodynamic response function (HRF). The prompts (indicated by an arrow) appeared on a computer screen 100 ms before scan offset and cued the participant to either produce a word (word prompt) or remain silent (blank prompt). The 1600-ms-long scans were separated by the 1600-ms-long silent periods permitting speaking and listening.

Participants were scanned in three functional runs, each lasting 9 min and comprising 168 trials (24 of each condition). Trials were presented in “blocks” of 7 trials: within each block, one trial of each of the seven conditions was presented. Across the whole experiment, every condition followed every condition (including itself) an approximately equal number of times. The stimuli for listen trials were taken from the recordings of the matched production trials in the preceding block, except for the first block, where the listen stimuli were taken from the production trials in the same block (i.e., production condition trials always preceded listen condition trials in this first block).

Each participant practiced 3 blocks of the experiment (i.e., 21 trials) in the scanner before scanning commenced. During the imaging session, behavior on each trial was monitored in real time by the experimenter (Z.Z.Z.), so that invalid trials (see below) could be excluded. Each participant’s vocal production and auditory feedback signals were segregated into two different channels and therefore monitored and recorded simultaneously.

Behavioral data analysis. Recordings from vocal production and auditory feedback on each trial were reviewed to ensure that invalid trials were identified and properly accounted for. Trials were considered invalid if (1) participants failed to produce, produced the incorrect word, or vocalized during listening/rest trials; or (2) auditory feedback was not triggered because of the very low level of the vocal production (no-trigger trials). Invalid trials were excluded from further analysis, as were runs in which >25% of trials within the same condition (i.e., 6 trials) were invalid. When, across all three runs, 25% or more of trials within the same condition (i.e., 18 trials or more) were invalid, data from that participant were eliminated.

Multivoxel pattern analysis. Functional images were preprocessed and analyzed using SPM5 (www.fil.ion.ucl.ac.uk/spm/) and a custom-made, modular toolbox implemented in an automatic analysis pipeline system

(<https://github.com/rhodricusack/automaticanalysis/wiki>). Data were realigned to the first true functional scan of each run (after two dummy scans were discarded) without further preprocessing (to preserve spatial-pattern information in native subject space). Realigned data for each subject were then entered into a single-subject GLM using an event-related analysis procedure (Josephs and Henson, 1999). For each GLM, we modeled three successive trials of the same condition in each run as one regressor. This amounts to temporal smoothing of the data, leading to a better signal-to-noise ratio and increased sensitivity of parameter estimates for the time series (Meyer et al., 2011). The modeling resulted in eight regressors associated with each experimental condition, for each of the three runs (i.e., 24 trials per condition/3 successive trials per regressor). In addition, invalid trials, if present, were modeled as a covariate of no interest for each run to reduce the error variance. We also included the six realignment parameters as regressors to ensure that variability resulting from head motion was accounted for. This model was convolved with the hemodynamic response function and then fitted to the MR time series in each voxel, resulting in parameter estimates (β) indexing the magnitude of response to the experimental conditions.

A searchlight analysis (Kriegeskorte et al., 2006), constrained to gray-matter voxels, was performed on the β images for each participant. The gray matter mask applied to the original whole-brain β images was segmented from each participant’s native-space structural image and then coregistered with the individual’s echo planar imaging space. We extracted, for each of the 24 regressors for each condition (eight in each of the three runs), the multivoxel pattern of voxel β values within each spherical searchlight of 4 mm radius (for all searchlights containing at least 30 gray matter voxels). Therefore, for each participant, each condition was associated with 24 multivoxel patterns in each sphere. Patterns were then compared with each other, within and across conditions, using Spearman correlation treating every voxel in the searchlight as a data

point (Haxby et al., 2001; Kriegeskorte et al., 2008). This led to the repeated measures of pattern similarity within a single condition and between each pair of experimental conditions across the experiment.

We then constructed, for each experimental contrast to be tested, a “similarity structure contrast matrix” that contained hypothesis-driven predictions regarding the relative magnitude of pattern correlations within and between conditions. Here, a similarity structure contrast matrix corresponded to a 6×6 (three feedback types within both production and listen conditions) correlation matrix representing the predicted pattern similarity between regressors within a single condition (on the diagonal) and between each pair of experimental conditions (in the off-diagonal cells).

The searchlight was centered on each voxel in turn (Kriegeskorte et al., 2006), and for each participant, a GLM was assessed for the center voxel within each searchlight, with the repeatedly measured pattern similarity (i.e., magnitude of pattern correlations) as the dependent variable and a similarity structure contrast matrix as the predictors. The resulting images of parameter estimates (β values), each corresponding to one of the experimental contrasts performed, were spatially transformed into MNI space (Mazziotta et al., 1995) using a nonlinear stereotaxic normalization procedure (Friston et al., 1995) and smoothed with an 8 mm FWHM Gaussian kernel, to compensate for anatomical variability across participants. Each set of contrast images were then entered into random-effects analyses (one-sample t tests) at the group level, comparing the mean parameter-estimate difference over subjects to zero. Clusters that survived the statistical threshold of $p < 0.05$ corrected for multiple comparisons over the whole brain using Gaussian random field theory (i.e., familywise error correction at the cluster level) (Worsley, 1996), were deemed significant.

Results

Behavioral data

Both real-time monitoring during scanning and inspection of recorded behavioral data indicated that participants followed the instructions in all trials. In production trials, they did not produce the incorrect word, or fail to produce any word, and in listening trials they did not produce speech. There were, however, some production trials in which the vocalization failed to trigger the delivery of auditory feedback (i.e., invalid trials). Based on the predefined elimination criteria, we identified four participants with elevated rates of such no-trigger, invalid, trials (i.e., $>25\%$ of trials in the same production condition across three runs), and these participants were excluded from further analysis. For the remaining 16 participants, the distribution of invalid no-trigger trials did not differ across sex of participant (6 females and 10 males; $\chi^2 = 1.182$, $p = 0.277$), runs ($\chi^2 = 5.788$, $p = 0.055$), or production conditions ($\chi^2 = 4.692$, $p = 0.096$). These invalid trials were modeled as a covariate of no interest in our imaging analysis.

Multivoxel pattern analysis (MVPA)

The MVPA was based on an assessment of the model fit between predicted and measured multivoxel pattern similarities within and between conditions. The resulting brain maps reflect the localization of regions in which multivoxel patterns were consistent with the expected pattern similarity for a given effect of

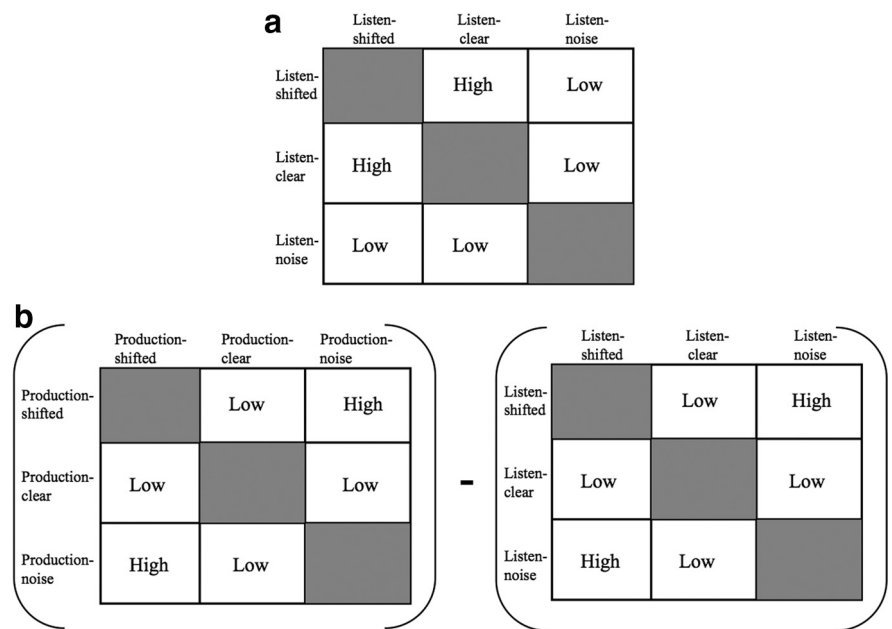


Figure 2. Similarity structure contrast matrices for the tested contrasts in our study. **a**, For speech-sensitive regions, we predicted greater similarity in multivoxel patterns between the two speech stimuli than between speech stimuli and noise during passive listening. **b**, For the regions sensitive to distorted auditory feedback, we predicted an interaction pattern between feedback types and production/listening: in other words, a greater similarity between the two types of distorted feedback than between either type of distorted feedback and normal feedback, during production but not during listening. This was tested by conducting a paired t test on the individually generated similarity maps from the minuend and subtrahend of the formulation shown here.

interest, or contrast. Anatomical structures were identified based on the LPBA40 probabilistic brain atlas (Shattuck et al., 2008).

Highlighting speech-sensitive regions

We reasoned that brain regions in which multivoxel patterns are consistently similar for speech stimuli must be involved in processing features characteristic of speech. Such regions will be expected to reveal multivoxel patterns that are as follows: (1) highly correlated between listening to clear speech (listen-clear) and listening to shifted speech (listen-shifted) conditions; and (2) less strongly correlated between listening to clear speech (listen-clear) and listening to signal-correlated noise (listen-noise), and between listening to shifted speech (listen-shifted) and listening to noise (listen-noise). These between-condition similarity predictions ensure that the multivoxel patterns contain information that is generalizable across different stimulus types that share a crucial “feature” (e.g., the presence of speech information), not just within a stimulus type. The predicted similarity structure among conditions (i.e., the similarity structure contrast matrix) is schematically presented in Figure 2a).

A one-group t test on the contrast images resulting from GLM evaluation of the “fit” between observed similarity and the predicted similarity structure contrast matrix revealed clusters localized to bilateral superior temporal gyrus (STG), extending anteriorly and ventrally into the superior temporal sulcus and middle temporal gyrus (MTG) (Table 1; Fig. 3). The observation of strong bilateral STG/MTG clusters in the neighborhood of putative speech-sensitive areas is consistent with a large body of literature on the perception of speech using a variety of stimuli/paradigms (for review, see Table 1 from Zheng et al., 2010). Our MVPA analysis yielded results that are highly compatible with the literature on this benchmark contrast compared with conventional univariate voxelwise analysis, attesting to the validity of our analytic framework.

Table 1. Speech-sensitive areas revealed by analysis of multivoxel patterns of brain activity

x	y	z	t value	Cluster size (no. of voxels)	Anatomical areas
−62	−32	−8	12.75	9605	Left STG/MTG
70	−22	−10	11.04	8431	Right STG/MTG

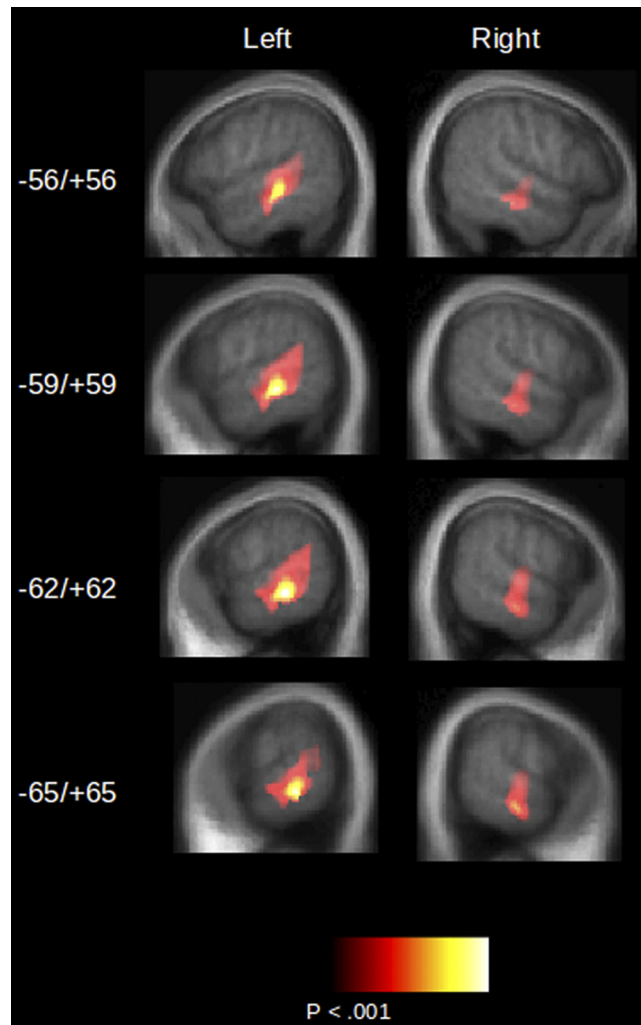


Figure 3. Speech-sensitive areas in which multivoxel patterns are more similar between listening to clear speech and listening to shifted speech, than between listening to either speech stimulus and listening to noise. Results are shown at $p < 0.001$ (familywise error rate corrected for multiple comparisons at the cluster level).

Highlighting networks that are differentially sensitive to auditory feedback error during vocalization, compared with during passive listening

The brain regions involved in auditory feedback error processing during vocalization should be sensitive to the discrepancy between articulation and its auditory concomitant. This requires that multivoxel patterns of activity in such regions be similar whether they are evoked during production with shifted speech feedback (production-shifted) or during production with masking noise feedback (production-noise). Furthermore, pattern similarity in error-sensitive areas should be greater between these two distorted-feedback conditions than between either of these conditions and production with clear speech feedback (production-clear). This similarity structure contrast matrix is depicted on the left side (minuend) of the symbolic formulation

Table 2. Areas in which multivoxel patterns of activity were more similar for production-shifted and production-noise, than for production-shifted and production-clear, or for production-noise and production-clear

x	y	z	t value	Cluster size (no. of voxels)	Anatomical areas
−34	−44	−24	5.86	902	Left cerebellum
42	−54	34	4.99	1118	Right AG
10	16	46	4.73	3180	Right SMA
30	−58	−30	4.59	1017	Right cerebellum

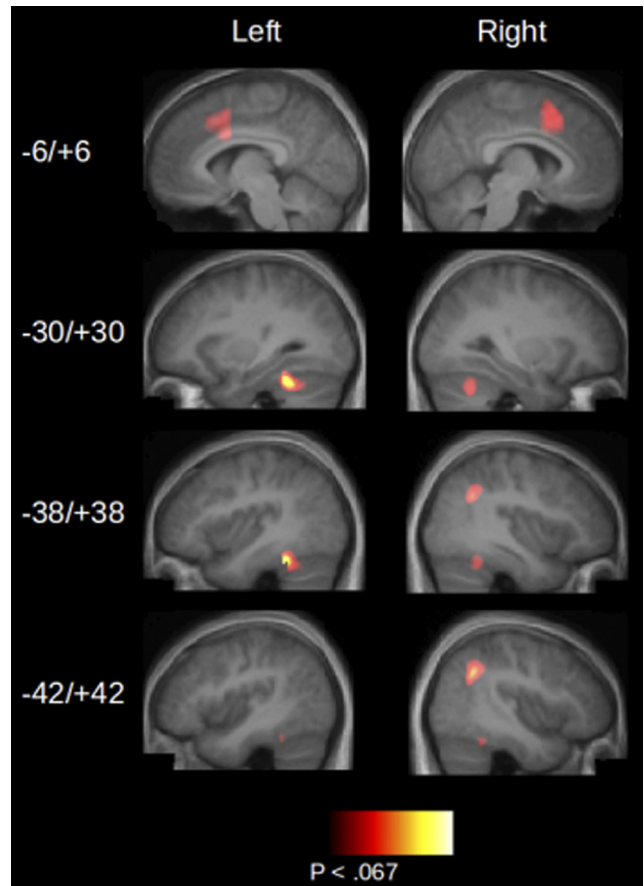


Figure 4. Error-coding areas in which multivoxel patterns are more similar between the two types of distorted feedback than between either distorted feedback and normal feedback, uniquely during production but not during passive listening.

in Figure 2*b*. Our MVPA analysis based on this production-related contrast matrix revealed significant clusters in the right posterior STG/inferior angular gyrus at the temporoparietal junction and in the right supplementary motor area (SMA). Additionally, there were marginally significant clusters located in the right cerebellum ($p = 0.053$) and in the left cerebellum ($p = 0.067$). The peaks of these cerebellum clusters survived a statistical threshold ($p < 0.05$ at the voxel level, false discovery rate corrected for multiple comparisons across the whole brain) (Genovese et al., 2002) (Table 2; Figure 4).

To explore whether the greater similarity between shifted speech and masking noise conditions was specific to production, we conducted the same MVPA analysis on the homologous listening conditions, as depicted in the listening-related contrast matrix on the right side (subtrahend) of the formulation in Figure 2*b*. However, no brain regions exhibited similarity patterns consistent with the prediction that listening to shifted speech (listen-shifted) and listening to noise (listen-noise) were significantly more similar to each other than they were to listening to clear speech (listen-clear).

The apparent difference between the similarity patterns elicited during production and those during listening (i.e., effectively the feedback type by production/listening interaction) was formally tested by comparing the contrast images for the MVPA analyses for production and listening using a paired *t* test: within SPM, paired *t* tests are directional (as set up, the paired *t* test would reveal regions in which the production-related contrast matrix is a significantly better fit to the data than is the listening-related contrast matrix, as shown in Fig. 2*b*). This “interaction” contrast revealed a distributed network of brain areas that partially overlapped with the network sensitive to distorted feedback during production. A significant interaction was observed in right inferior angular gyrus, right SMA, and bilateral cerebellum, as well as in a number of additional areas, including bilateral STG/MTG, extending from the most posterior regions in the STG to the anterior part of the MTG, and bilateral inferior frontal gyri (IFG) extending dorsally toward the precentral gyrus (Fig. 5).

We refer to the production and listening contrast matrices depicted in Figure 2*b* as production and listening “simple effects” by analogy with conventional ANOVA. The significant feedback type by production/listening interaction could arise either because the predicted similarity structure is a better fit with observed similarities during production than during listening or because the inverse of the predicted similarity structure fits the observed similarity pattern during listening better than the predicted similarity structure fits the production data. We refer to the inverse of the predicted similarity structure contrast matrix as the “inverse simple effect.” The inverse listening simple effect highlights regions in which multivoxel patterns of activity elicited by listen-shifted or listen-noise are more similar to the pattern elicited by listen-clear than they are to each other (i.e., in Fig. 2*b*; the labels in the similarity structure contrast matrix on the right hand side would flip such that “high” becomes “low” and vice versa).

To determine the relative contributions of the production and inverse listening simple effects to the observed interaction, we show the interaction inclusively masked by both effects in Figure 5. The interaction masked by the production simple effect yielded regions, including right angular gyrus, right SMA, and bilateral cerebellum, as shown in Figure 5 (cyan). The interaction masked by the inverse listening simple effect yielded regions including bilateral STG/MTG and a small portion of the left precentral gyrus, as shown in Figure 5 (magenta). A breakdown of the inverse listening simple effect, comparing the similarity of listen-clear to either listen-shifted or listen-noise separately, indicates that similarity between listen-clear and listen-shifted was driving the patterns observed in bilateral STG/MTG and left precentral gyrus (Fig. 5, magenta). This is not surprising because both of these conditions are speech-like, unlike the auditory stimulus in the listen-noise condition, and the observed brain regions are known to be speech sensitive (e.g., Binder et al., 2000; Wilson et al., 2004; Obleser et al., 2007).

The bilateral IFG regions that were observed in the interaction did not seem to arise either because of the production simple effect or because of the inverse listening simple effect alone but appear to reflect both simple effects.

Assessing pattern similarity among conditions reveals functionally differentiated networks

We assessed functional specificity of the brain networks observed in the feedback type by production/listening interaction (Fig. 5, magenta, cyan, and blue), by exploring whether these networks generated differentiable response profiles across conditions. This

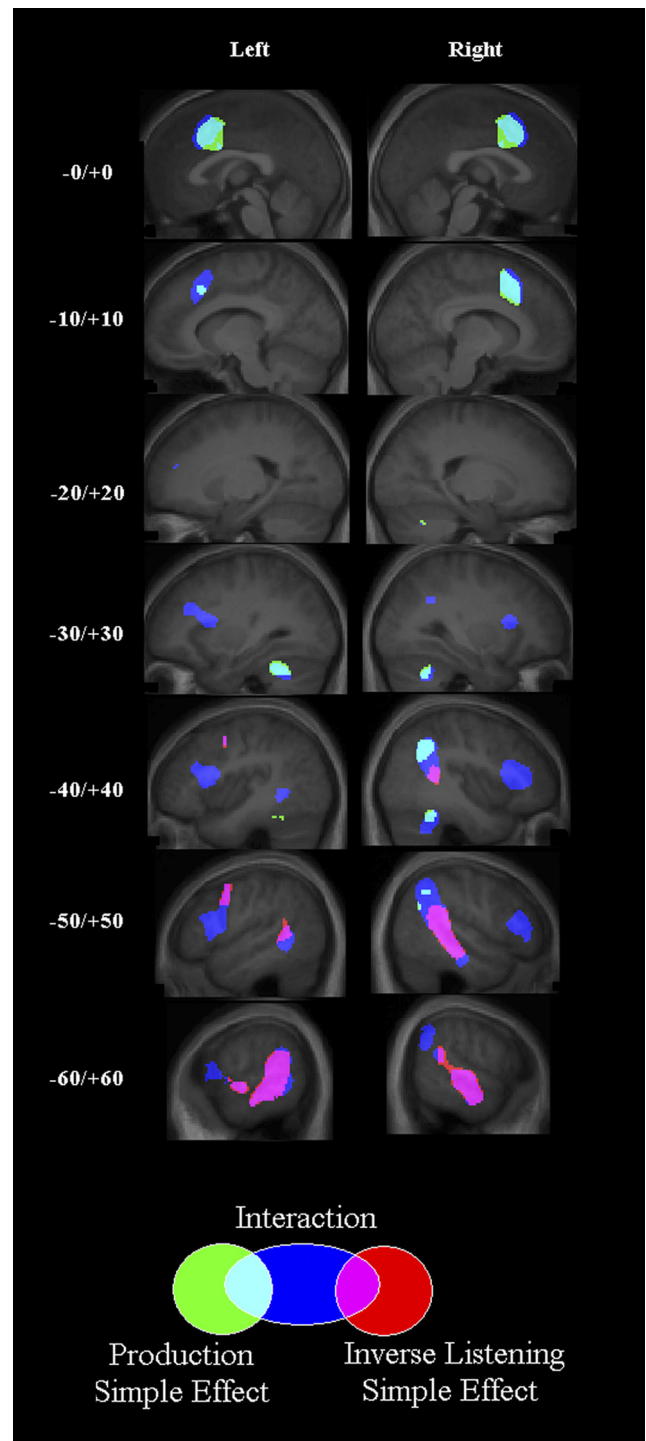


Figure 5. Brain areas in which multivoxel patterns exhibited a feedback type by production/listening interaction. This interaction was inclusively masked by production simple effect and by inverse listening simple effect to demonstrate their relative contributions: cyan represents predominantly driven by production simple effect; magenta, predominantly driven by inverse listening simple effect; blue, a combination of the two (blue).

was done by creating a mask for the volume of significant voxels within each network and then extracting the mean correlation coefficients for the six between-condition correlation pairs (i.e., three pairs, clear/shift, shift/noise, and clear/noise, for both production and listening). A repeated-measures ANOVA with networks (3 levels) and condition pairs (6 levels) as within-subject factors indicated that there was a significant interaction between

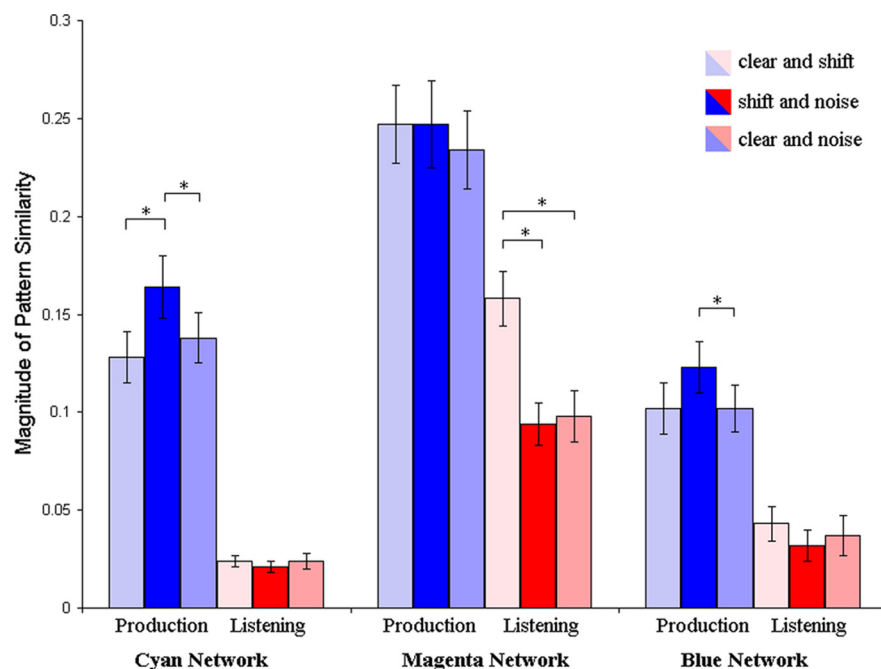


Figure 6. The magnitudes of the six between-condition pattern correlations are plotted for three functional networks. For each network, three pairs of comparisons are shown for production (shades of blue on the left) and listening (shades of red on the right). Vertical bars indicate SE. *Significant difference between the means ($p < 0.05$).

networks and condition pairs: $F_{(6,10)} = 9.67$, $p = 0.006$ (Fig. 6). Three separate ANOVAs, each testing one pair of networks, confirmed that the patterns of correlation across the 6 condition pairs differed significantly among networks: (network by condition pair interaction: cyan vs magenta: $F_{(5,11)} = 26.34$, $p < 0.001$; cyan vs blue: $F_{(5,11)} = 4.58$, $p = 0.017$; magenta vs blue: $F_{(5,11)} = 18.90$, $p < 0.001$). *Post hoc* comparisons revealed that for the cyan network (i.e., production simple effect), production-shift/noise pattern correlations were significantly stronger than either production-shift/clear correlations ($p = 0.001$) or production-clear/noise correlations ($p = 0.003$). For the magenta network (i.e., inverse listening simple effect), listen-clear/shift pattern correlations were significantly stronger than either listen-shift/noise correlations ($p < 0.001$) or listen-clear/noise correlations ($p < 0.001$). For the blue network (i.e., showing neither simple effect clearly but sharing features of both), three of the four pairwise comparisons from the two simple effects demonstrated significant or nearly significant differences: production-shift/noise correlations being stronger than production-clear/shift correlations ($p = 0.052$) or production-clear/noise correlations ($p = 0.007$); listen-clear/shift correlations being stronger than listen-shift/noise correlations ($p = 0.054$). These results indicate that the feedback type by production/listening interaction arises from three distinct profiles of between-condition pattern similarity, suggesting that the brain networks exhibiting these distinct profiles are functionally dissociable.

Discussion

In the present study, we combined fMRI and an MVPA framework to explore the functional architecture subserving auditory feedback processing of speech and its role in the control of speech. Whole-brain analysis of multivoxel neural pattern similarity revealed three functionally differentiable networks that exhibited different patterns of sensitivity to auditory input across production and listening conditions. The distinct patterns of sensitivity, presumably reflecting the operation of functionally spe-

cialized networks (van de Ven et al., 2009), suggest a distributed neural architecture supporting sensorimotor processes essential to control of speech production.

One specialized network (cyan network), including bilateral cerebellum, right angular gyrus, and right SMA, yielded patterns of activity that were similar for the two acoustically different, distorted-feedback conditions but were markedly less similar for clear speech feedback compared with either distorted-feedback condition. Furthermore, this profile was observed uniquely during articulation but not during listening, a pattern that would be consistent with encoding an error signal during talking. That the cerebellum is part of this network is not surprising: other studies have suggested that it integrates sensory and motor inputs to control the precision and timing of movement (Jacobson et al., 2008; Dean et al., 2010). Clinical patients with cerebellar lesions consistently show uncoordinated, disruptive, or impaired motion (Diener and Dichgans, 1992; Baier et al., 2009), largely because of poor utilization of sensory-feedback information. The

cerebellum maintains an internal model that is capable of adaptively updating motor commands to achieve an intended motor output (Wolpert et al., 1998; Blakemore et al., 2001; Ito, 2008). This function is highly relevant for speech motor control (Riecker et al., 2006; Callan et al., 2007; Ackermann, 2008), particularly when, as in our case, the expected and actual sensory consequences of the motor output do not match.

The right angular gyrus has been implicated in monitoring one's own visuomotor movements, acting as a high-level motor control center in detecting mismatch between predicted and perceived sensory outcome (Sirigu et al., 2004). Farrer et al. (2008) observed significant neural activation in the right angular gyrus when the correspondence between the intended and perceived sensory consequences of a motor movement was manipulated by way of a sensory feedback delay. The magnitude of brain activation in this area also correlated with the degree of the discrepancy (Farrer et al., 2003), indicating a neural representation of sensory feedback error.

The final region observed to be part of this network was the SMA bordering on dorsal anterior cingulate cortex. Given the relatively rostral location of the SMA peak in the current study, it primarily encompasses the pre-SMA subdivision (Picard and Strick, 1996; Tanji, 1996). The pre-SMA interconnects with prefrontal cortex (Bates and Goldman-Rakic, 1993; Luppino et al., 1993) and is known to be involved in motor preparation and planning (Hikosaka et al., 1996), general sensorimotor integration (Kurata et al., 2000), and updating of motor commands and plans (Matsuzaka and Tanji, 1996; Shima et al., 1996). The dorsal anterior cingulate cortex appears to be involved in error detection and conflict monitoring (Garavan et al., 2003) and is critical to the processing of error-related responses (Hester et al., 2004). In sum, all of the areas highlighted in this network are highly compatible with a function of encoding an articulatory error signal and adjustment of subsequent motor commands.

The second specialized network (magenta network) includes bilateral anterolateral STG and left precentral gyrus. In these regions, multivoxel patterns were more similar between clear speech feedback and formant-shifted speech feedback than between either of these conditions and the noise-feedback condition. Furthermore, this speech-sensitive pattern was present only during listening and not during production. The observed network has consistently been implicated in speech processing (Scott et al., 2000; e.g., Binder et al., 2000; Narain et al., 2003; Wilson et al., 2004; Liebenthal et al., 2005; Obleser et al., 2007), but here we demonstrate that speech sensitivity in this network is not evident during production. Our results show that these regions are not functionally restricted to speech perception; they appear to change their function during production, such that sensitivity to heard speech is attenuated. This is consistent with previous literature demonstrating a role of bilateral STG in auditory feedback control of speech (Fu et al., 2006; Christoffels et al., 2007; Tourville et al., 2008; Zheng et al., 2010).

A final network (blue network), localized to bilateral IFG, appears to share characteristics with both of the other two networks. The multivoxel patterns in these regions were more similar for the two distorted-feedback conditions during production and demonstrated speech sensitivity during listening, although this pattern was not significant. Previous work implicates IFG regions in visuomotor interactions (Johnson-Frey et al., 2003), motor execution based on observation (Iacoboni et al., 1999), and sensorimotor integration in motor action (Parsons et al., 2005). Here these frontal regions may be involved in using incoming auditory information to modify how articulatory gestures are programmed.

We have extended previous correlation-based MVPA approaches (Haxby et al., 2001; O'Toole et al., 2005; Downing et al., 2007; Williams et al., 2007; Kay et al., 2008; Kriegeskorte et al., 2008; Peelen et al., 2009; Stokes et al., 2009), by incorporating hypothesis-driven predictions (Linke et al., 2011). These predictions are formulated as individual contrast-like matrices and tested against the measured correlations of multivoxel patterns through GLMs, leading to model-based estimates of within- and between-condition similarity. The use of GLMs circumvents the choice of complex algorithms (Cox and Savoy, 2003; e.g., Carlson et al., 2003; Mitchell et al., 2004; Haynes and Rees, 2005; Kamitani and Tong, 2005; Kriegeskorte et al., 2006; Pessoa and Padmala, 2006; Serences and Boynton, 2007; Friston et al., 2008) for pattern classification and renders the evaluation of experimentally relevant hypotheses intuitive. The methodological simplicity and conceptual clarity ensure that our method can be readily adapted to other studies to draw inferences about the functional architecture of brain networks.

It is important to acknowledge that, in the context of auditory feedback processing, experimental designs that are optimal to observe behavioral effects are not necessarily optimal for fMRI data acquisition. In the current design, the randomization required for optimal imaging eliminates the measurable behavioral compensation that would normally build up over many successive distorted-feedback trials. However, we did observe statistically reliable, condition-dependent brain responses, even when the overt behavioral changes were not evident. We think that such brain responses are interpretable because one of the strengths of neuroimaging as an experimental technique is that overt behavior is not always necessary (Owen and Coleman, 2008; Coleman et al., 2009; Monti et al., 2010; Cruse et al., 2011; see, for example, work in which neuroimaging is used to probe

for awareness in behaviorally unresponsive patients diagnosed in vegetative state).

We use the term “speech motor control” broadly, as it is commonly used in the field (e.g., Kent, 2000) to refer to all the processes from the intention to act, including those involved in generation of plans and initiation of movement, to feedback error monitoring and processing, to adjust the internal model. Our design was optimized to detect the rapid, immediate, neural consequences related to error detection and processing (e.g., Houde et al., 2002) and not for measurement of altered articulation. The possibility that we are merely tapping into a generic error-detection system (e.g., Kerns et al., 2004) rather than into a speech-specific one seems unlikely because of the extensive network, consistent with speech motor control observed in the present study. The possibility that the full set of processes related to speech motor control were not elicited by our approach also seems unlikely given the relatively automatic nature of the compensation in auditory speech perturbations (Munhall et al., 2009).

Computational models of speech motor control, including the DIVA (Guenther et al., 2006; Golfingopoulos et al., 2010) and State Feedback Control (Hickok et al., 2011; Houde and Nagarajan, 2011), are embodied within a large network of anatomical regions supporting feedforward and feedback control of vocal motor production. The feedback control systems in both of these models are assumed to be anatomically constrained and functionally unitary (e.g., Tourville et al., 2008; Golfingopoulos et al., 2010). In contrast, we demonstrate here that a subset of such control systems cover widely distributed brain areas and include functionally differentiated components. One distributed network is directly involved in capturing a speech error signal resulting from the disparity between what is spoken and what is heard. This network might be particularly important for detection of errors during speech movements. Another network appears to alter its function in the context of listening compared with the context of production, such that sensitivity to distorted feedback as an auditory concomitant of speech is accentuated. A third network appears to capture aspects of the previous two networks and may be involved in programming motor output on the basis of sensory information. These functionally heterogeneous networks operate in concert to support auditory feedback processing of speech, reflecting the complexity and intricacy of the neural processes underlying speech motor control.

References

- Ackermann H (2008) Cerebellar contributions to speech production and speech perception: psycholinguistic and neurobiological perspectives. *Trends Neurosci* 31:265–272. [CrossRef Medline](#)
- Baier B, Stoeter P, Dieterich M (2009) Anatomical correlates of ocular motor deficits in cerebellar lesions. *Brain* 132:2114–2124. [CrossRef Medline](#)
- Bates JF, Goldman-Rakic PS (1993) Prefrontal connections of medial motor areas in the rhesus monkey. *J Comp Neurol* 336:211–228. [CrossRef Medline](#)
- Behroozmand R, Larson CR (2011) Error-dependent modulation of speech-induced auditory suppression for pitch-shifted voice feedback. *BMC Neurosci* 12:54. [CrossRef Medline](#)
- Behroozmand R, Korzyukov O, Larson CR (2012) ERP correlates of pitch error detection in complex tone and voice auditory feedback with missing fundamental. *Brain Res* 1448:89–100. [CrossRef Medline](#)
- Binder JR, Frost JA, Hammeke TA, Bellgowan PS, Springer JA, Kaufman JN, Possing ET (2000) Human temporal lobe activation by speech and non-speech sounds. *Cereb Cortex* 10:512–528. [CrossRef Medline](#)
- Blakemore SJ, Frith CD, Wolpert DM (2001) The cerebellum is involved in predicting the sensory consequences of action. *Neuroreport* 12:1879–1884. [CrossRef Medline](#)

- Burnett TA, Freedland MB, Larson CR, Hain TC (1998) Voice F0 responses to manipulations in pitch feedback. *J Acoust Soc Am* 103:3153–3161. [CrossRef Medline](#)
- Callan DE, Kawato M, Parsons L, Turner R (2007) Speech and song: the role of the cerebellum. *Cerebellum* 6:321–327. [CrossRef Medline](#)
- Carlson TA, Schrater P, He S (2003) Patterns of activity in the categorical representation of objects. *J Cogn Neurosci* 15:704–717. [CrossRef Medline](#)
- Christoffels IK, Formisano E, Schiller NO (2007) Neural correlates of verbal feedback processing: an fMRI study employing overt speech. *Hum Brain Mapp* 28:868–879. [CrossRef Medline](#)
- Christoffels IK, van de Ven V, Waldorp LJ, Formisano E, Schiller NO (2011) The sensory consequences of speaking: parametric neural cancellation during speech in auditory cortex. *PLoS One* 6:e18307. [CrossRef Medline](#)
- Coleman MR, Davis MH, Rodd JM, Robson T, Ali A, Owen AM, Pickard JD (2009) Towards the routine use of brain imaging to aid the clinical diagnosis of disorders of consciousness. *Brain* 132:2541–2552. [CrossRef Medline](#)
- Cox DD, Savoy RL (2003) Functional magnetic resonance imaging (fMRI) “brain reading”: detecting and classifying distributed patterns of fMRI activity in human visual cortex. *Neuroimage* 19:261–270. [CrossRef Medline](#)
- Cruse D, Chennu S, Chatelle C, Bekinschtein TA, Fernández-Espejo D, Pickard JD, Laureys S, Owen AM (2011) Bedside detection of awareness in the vegetative state: a cohort study. *Lancet* 378:2088–2094. [Medline](#)
- Dean P, Porrill J, Ekerot CF, Jörntell H (2010) The cerebellar microcircuit as an adaptive filter: experimental and computational evidence. *Nat Rev Neurosci* 11:30–43. [CrossRef Medline](#)
- Diener HC, Dichgans J (1992) Pathophysiology of cerebellar ataxia. *Mov Disord* 7:95–109. [CrossRef Medline](#)
- Donath TM, Natke U, Kalveram KT (2002) Effects of frequency-shifted auditory feedback: a preliminary study F0 contours in syllables. *J Acoust Soc Am* 111:357–366. [CrossRef Medline](#)
- Downing PE, Wiggett AJ, Peelen MV (2007) Functional magnetic resonance imaging investigation of overlapping lateral occipitotemporal activations using multi-voxel pattern analysis. *J Neurosci* 27:226–233. [CrossRef Medline](#)
- Eliades SJ, Wang X (2008) Neural substrates of vocalization feedback monitoring in primate auditory cortex. *Nature* 453:1102–1106. [CrossRef Medline](#)
- Farrer C, Franck N, Georgieff N, Frith CD, Decety J, Jeannerod M (2003) Modulating the experience of agency: a positron emission tomography study. *Neuroimage* 18:324–333. [CrossRef Medline](#)
- Farrer C, Frey SH, Van Horn JD, Tunik E, Turk D, Inati S, Grafton ST (2008) The angular gyrus computes action awareness representations. *Cereb Cortex* 18:254–261. [CrossRef Medline](#)
- Friston KJ, Ashburner J, Frith CD, Poline JB, Heather JD, Frakowiak RSJ (1995) Spatial registration and normalization of images. *Hum Brain Mapp* 3:165–189. [CrossRef](#)
- Friston K, Chu C, Mourão-Miranda J, Hulme O, Rees G, Penny W, Ashburner J (2008) Bayesian decoding of brain images. *Neuroimage* 39:181–205. [CrossRef Medline](#)
- Fu CH, Vythelingum GN, Brammer MJ, Williams SC, Amaro E Jr, Andrew CM, Yáñez L, van Haren NE, Matsumoto K, McGuire PK (2006) An fMRI study of verbal self-monitoring: neural correlates of auditory verbal feedback. *Cereb Cortex* 16:969–977. [CrossRef Medline](#)
- Garavan H, Ross TJ, Kaufman J, Stein EA (2003) A midline rostral-caudal axis for error processing and response conflict monitoring. *Neuroimage* 20:1132–1139. [CrossRef Medline](#)
- Genovese CR, Lazar NA, Nichols T (2002) Thresholding of statistical maps in functional neuroimaging using the false discovery rate. *Neuroimage* 15:870–878. [CrossRef Medline](#)
- Golfopoulos E, Tourville JA, Guenther FH (2010) The integration of large-scale neural network modeling and functional brain imaging in speech motor control. *Neuroimage* 52:862–874. [CrossRef Medline](#)
- Guenther FH, Ghosh SS, Tourville JA (2006) Neural modeling and imaging of the cortical interactions underlying syllable production. *Brain Lang* 96:280–301. [CrossRef Medline](#)
- Hashimoto Y, Sakai KL (2003) Brain activations during conscious self-monitoring of speech production with delayed auditory feedback: an fMRI study. *Hum Brain Mapp* 20:22–28. [CrossRef Medline](#)
- Haxby JV, Gobbini MI, Furey ML, Ishai A, Schouten JL, Pietrini P (2001) Distributed and overlapping representations of faces and objects in ventral temporal cortex. *Science* 293:2425–2430. [CrossRef Medline](#)
- Haynes JD, Rees G (2005) Predicting the orientation of invisible stimuli from activity in human primary visual cortex. *Nat Neurosci* 8:686–691. [CrossRef Medline](#)
- Haynes JD, Rees G (2006) Decoding mental states from brain activity in humans. *Nat Rev Neurosci* 7:523–534. [CrossRef Medline](#)
- Heinks-Maldonado TH, Mathalon DH, Gray M, Ford JM (2005) Fine-tuning of auditory cortex during speech production. *Psychophysiology* 42:180–190. [CrossRef Medline](#)
- Hester R, Fassbender C, Garavan H (2004) Individual differences in error processing: a review and reanalysis of three event-related fMRI studies using the GO/NOGO task. *Cereb Cortex* 14:986–994. [CrossRef Medline](#)
- Hickok G, Houde J, Rong F (2011) Sensorimotor integration in speech processing: computational basis and neural organization. *Neuron* 69:407–422. [CrossRef Medline](#)
- Hikosaka O, Sakai K, Miyauchi S, Takino R, Sasaki Y, Pütz B (1996) Activation of human presupplementary motor area in learning of sequential procedures: a functional MRI study. *J Neurophysiol* 76:617–621. [Medline](#)
- Houde JF, Jordan MI (1998) Sensorimotor adaptation in speech production. *Science* 279:1213–1216. [CrossRef Medline](#)
- Houde JF, Nagarajan SS (2011) Speech production as state feedback control. *Front Hum Neurosci* 5:82. [CrossRef Medline](#)
- Houde JF, Nagarajan SS, Sekihara K, Merzenich MM (2002) Modulation of the auditory cortex during speech: an MEG study. *J Cogn Neurosci* 14:1125–1138. [CrossRef Medline](#)
- Iacoboni M, Woods RP, Brass M, Bekkering H, Mazziotta JC, Rizzolatti G (1999) Cortical mechanisms of human imitation. *Science* 286:2626–2628. [Medline](#)
- Ito M (2008) Control of mental activities by internal models in the cerebellum. *Nat Rev Neurosci* 9:304–313. [CrossRef Medline](#)
- Jacobson GA, Rokni D, Yarom Y (2008) A model of the olivo-cerebellar system as a temporal pattern generator. *Trends Neurosci* 31:617–625. [CrossRef Medline](#)
- Johnson-Frey SH, Maloof FR, Newman-Norlund R, Farrer C, Inati S, Grafton ST (2003) Actions or hand-object interactions? Human inferior frontal cortex and action observation. *Neuron* 39:1053–1058. [CrossRef Medline](#)
- Josephs O, Henson RN (1999) Event-related functional magnetic resonance imaging: modeling, inference and optimization. *Philos Trans R Soc Lond B Biol Sci* 354:1215–1228. [CrossRef Medline](#)
- Kamitani Y, Tong F (2005) Decoding the visual and subjective contents of the human brain. *Nat Neurosci* 8:679–685. [CrossRef Medline](#)
- Kay KN, Naselaris T, Prenger RJ, Gallant JL (2008) Identifying natural images from human brain activity. *Nature* 452:352–355. [CrossRef Medline](#)
- Kent RD (2000) Research on speech motor control and its disorders: a review and prospective. *J Commun Disord* 33:391–427; quiz 428. [CrossRef Medline](#)
- Kerns JG, Cohen JD, MacDonald AW 3rd, Cho RY, Stenger VA, Carter CS (2004) Anterior cingulate conflict monitoring and adjustments in control. *Science* 303:1023–1026. [CrossRef Medline](#)
- Kriegeskorte N (2011) Pattern-information analysis: from stimulus decoding to computational-model testing. *Neuroimage* 56:411–421. [CrossRef Medline](#)
- Kriegeskorte N, Goebel R, Bandettini P (2006) Information-based functional brain mapping. *Proc Natl Acad Sci U S A* 103:3863–3868. [CrossRef Medline](#)
- Kriegeskorte N, Mur M, Bandettini P (2008) Representational similarity analysis: connecting the branches of systems neuroscience. *Front Syst Neurosci* 2:4. [CrossRef Medline](#)
- Kurata K, Tsuji T, Naraki S, Seino M, Abe Y (2000) Activation of the dorsal premotor cortex and pre-supplementary motor area of humans during an auditory conditional motor task. *J Neurophysiol* 84:1667–1672. [Medline](#)
- Levelt WJ (1983) Monitoring and self-repair in speech. *Cognition* 14:41–104. [CrossRef Medline](#)
- Levelt WJ, Roelofs A, Meyer AS (1999) A theory of lexical access in speech production. *Behav Brain Sci* 22:1–38. [Medline](#)
- Liebsenthal E, Binder JR, Spitzer SM, Possing ET, Medler DA (2005) Neural substrates of phonemic perception. *Cereb Cortex* 15:1621–1631. [CrossRef Medline](#)
- Linke AC, Vicente-Grabovetsky A, Cusack R (2011) Stimulus-specific suppression preserves information in auditory short-term memory. *Proc Natl Acad Sci U S A* 108:12961–12966. [CrossRef Medline](#)

- Luppino G, Matelli M, Camarda R, Rizzolatti G (1993) Corticocortical connections of area F3 (SMA-proper) and area F6 (pre-SMA) in the macaque monkey. *J Comp Neurol* 338:114–140. [CrossRef Medline](#)
- MacDonald EN, Goldberg R, Munhall KG (2010) Compensations in response to real-time formant perturbations of different magnitudes. *J Acoust Soc Am* 127:1059–1068. [CrossRef Medline](#)
- MacDonald EN, Purcell DW, Munhall KG (2011) Probing the independence of formant control using altered auditory feedback. *J Acoust Soc Am* 129:955–965. [CrossRef Medline](#)
- Matsuzaka Y, Tanji J (1996) Changing directions of forthcoming arm movements: neuronal activity in the presupplementary and supplementary motor area of monkey cerebral cortex. *J Neurophysiol* 76:2327–2342. [Medline](#)
- Mazziotta JC, Toga AW, Evans A, Fox P, Lancaster J (1995) A probabilistic atlas of the human brain: theory and rationale for its development. The international Consortium for Brain Mapping (ICBM). *Neuroimage* 2:89–101. [CrossRef Medline](#)
- McGuire PK, Silbersweig DA, Frith CD (1996) Functional neuroanatomy of verbal self-monitoring. *Brain* 119:907–917. [CrossRef Medline](#)
- Meyer K, Kaplan JT, Essex R, Damasio H, Damasio A (2011) Seeing touch is correlated with content-specific activity in primary somatosensory cortex. *Cereb Cortex* 21:2113–2121. [CrossRef Medline](#)
- Mitchell TM, Hutchinson R, Niculescu RS, Pereira F, Wang X (2004) Learning to decode cognitive states from brain images. *Mach Learn* 57:145–175. [CrossRef](#)
- Monti MM, Vanhaudenhuyse A, Coleman MR, Boly M, Pickard JD, Tshibanda L, Owen AM, Laureys S (2010) Willful modulation of brain activity in disorders of consciousness. *N Engl J Med* 362:579–589. [CrossRef Medline](#)
- Munhall KG, MacDonald EN, Byrne SK, Johnsrude I (2009) Talkers alter vowel production in response to real-time formant perturbation even when instructed not to compensate. *J Acoust Soc Am* 125:384–390. [CrossRef Medline](#)
- Narain C, Scott SK, Wise RJ, Rosen S, Leff A, Iversen SD, Matthews PM (2003) Defining a left-lateralized response specific to intelligible speech using fMRI. *Cereb Cortex* 13:1362–1368. [CrossRef Medline](#)
- Norman KA, Polyn SM, Detre GJ, Haxby JV (2006) Beyond mind-reading: multi-vowel pattern analysis of fMRI data. *Trends Cogn Sci* 10:424–430. [CrossRef Medline](#)
- Obleser J, Zimmermann J, Van Meter J, Rauschecker JP (2007) Multiple stages of auditory speech perception reflected in event-related fMRI. *Cereb Cortex* 17:2251–2257. [CrossRef Medline](#)
- O'Toole AJ, Jiang F, Abdi H, Haxby JV (2005) Partially distributed representations of objects and faces in ventral temporal cortex. *J Cogn Neurosci* 17:580–590. [CrossRef Medline](#)
- Orfanidis SJ (1988) Optimum signal processing, an introduction. New York: MacMillan.
- Orfanidou E, Marslen-Wilson WD, Davis MH (2006) Neural response suppression predicts repetition priming of spoken words and pseudowords. *J Cogn Neurosci* 18:1237–1252. [CrossRef Medline](#)
- Owen AM, Coleman MR (2008) Functional neuroimaging of the vegetative state. *Nat Rev Neurosci* 9:235–243. [CrossRef Medline](#)
- Parsons LM, Sergent J, Hodges DA, Fox PT (2005) The brain basis of piano performance. *Neuropsychologia* 43:199–215. [CrossRef Medline](#)
- Paus T, Perry DW, Zatorre RJ, Worsley KJ, Evans AC (1996) Modulation of cerebral blood flow in the human auditory cortex during speech: role of motor-to-sensory discharges. *Eur J Neurosci* 8:2236–2246. [CrossRef Medline](#)
- Peelen MV, Fei-Fei L, Kastner S (2009) Neural mechanisms of rapid natural scene categorization in human visual cortex. *Nature* 460:94–97. [CrossRef Medline](#)
- Pessoa L, Padmala S (2007) Decoding near-threshold perception of fear from distributed single-trial brain activation. *Cereb Cortex* 17:691–701. [CrossRef Medline](#)
- Picard N, Strick PL (1996) Motor areas of the medial wall: a review of their location and functional activation. *Cereb Cortex* 6:342–353. [CrossRef Medline](#)
- Postma A (2000) Detection of errors during speech production: a review of speech monitoring models. *Cognition* 77:97–132. [CrossRef Medline](#)
- Purcell DW, Munhall KG (2006a) Compensation following real-time manipulation of formants in isolated vowels. *J Acoust Soc Am* 119:2288–2297. [CrossRef Medline](#)
- Purcell DW, Munhall KG (2006b) Adaptive control of vowel formant frequency: evidence from real-time formant manipulation. *J Acoust Soc Am* 120:966–977. [CrossRef Medline](#)
- Riecker A, Kassubek J, Gröschel K, Grodd W, Ackermann H (2006) The cerebral control of speech tempo: opposite relationship between speaking rate and BOLD signal changes at striatal and cerebellar structures. *Neuroimage* 29:46–53. [CrossRef Medline](#)
- Schroeder MR (1968) Reference signal for signal quality studies. *J Acoust Soc Am* 44:1735–1736. [CrossRef](#)
- Scott SK, Blank CC, Rosen S, Wise RJ (2000) Identification of a pathway for intelligible speech in the left temporal lobe. *Brain* 123:2400–2406. [CrossRef Medline](#)
- Serences JT, Boynton GM (2007) The representation of behavioral choice for motion in human visual cortex. *J Neurosci* 27:12893–12899. [CrossRef Medline](#)
- Shattuck DW, Mirza M, Adisetiyo V, Hojatkashani C, Salamon G, Narr KL, Poldrack RA, Bilder RM, Toga AW (2008) Construction of a 3D probabilistic atlas of human cortical structures. *Neuroimage* 39:1064–1080. [CrossRef Medline](#)
- Shima K, Mushiake H, Saito N, Tanji J (1996) Role for cells in the presupplementary motor area in updating motor plans. *Proc Natl Acad Sci U S A* 93:8694–8698. [CrossRef Medline](#)
- Sirigu A, Daprati E, Ciancia S, Giraux P, Nighoghossian N, Posada A, Haggard P (2004) Altered awareness of voluntary action after damage to the parietal cortex. *Nat Neurosci* 7:80–84. [CrossRef Medline](#)
- Stokes M, Thompson R, Cusack R, Duncan J (2009) Top-down activation of shape-specific population codes in visual cortex during mental imagery. *J Neurosci* 29:1565–1572. [CrossRef Medline](#)
- Tanji J (1996) New concepts of the supplementary motor area. *Curr Opin Neurobiol* 6:782–787. [CrossRef Medline](#)
- Tourville JA, Reilly KJ, Guenther FH (2008) Neural mechanisms underlying auditory feedback control of speech. *Neuroimage* 39:1429–1443. [CrossRef Medline](#)
- van de Ven V, Esposito F, Christoffels IK (2009) Neural network of speech monitoring overlaps with overt speech production and comprehension networks: a sequential spatial and temporal ICA study. *Neuroimage* 47:1982–1991. [CrossRef Medline](#)
- Villacorta VM, Perkell JS, Guenther FH (2007) Sensorimotor adaptation to feedback perturbations of vowel acoustics and its relation to perception. *J Acoust Soc Am* 122:2306–2319. [CrossRef Medline](#)
- Williams MA, Dang S, Kanwisher NG (2007) Only some spatial patterns of fMRI response are read out in task performance. *Nat Neurosci* 10:685–686. [CrossRef Medline](#)
- Wilson SM, Saygin AP, Sereno MI, Iacoboni M (2004) Listening to speech activates motor areas involved in speech production. *Nat Neurosci* 7:701–702. [CrossRef Medline](#)
- Wolpert DM, Miall RC, Kawato M (1998) Internal models in the cerebellum. *Trends Cogn Sci* 2:338–347. [CrossRef Medline](#)
- Worsley KJ, Marrett S, Neelin P, Vandal AC, Friston KJ, Evans AC (1996) A unified statistical approach for determining significant signals in images of cerebral activation. *Hum Brain Mapp* 4:58–73. [CrossRef Medline](#)
- Zheng ZZ, Munhall KG, Johnsrude IS (2010) Functional overlap between regions involved in speech perception and in monitoring one's own voice during speech production. *J Cogn Neurosci* 22:1770–1781. [CrossRef Medline](#)